

AD-A103 010

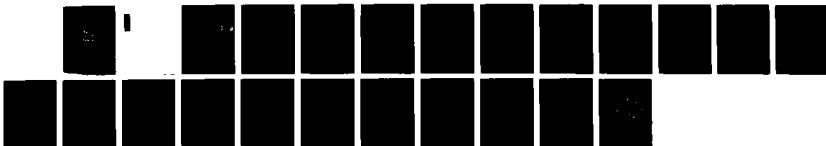
ESTIMATION IN PARAMETRIC MIXTURE FAMILIES(U) STANFORD  
UNIV CA DEPT OF STATISTICS A E GELFAND 18 AUG 87  
TR-395 N00014-86-K-0156

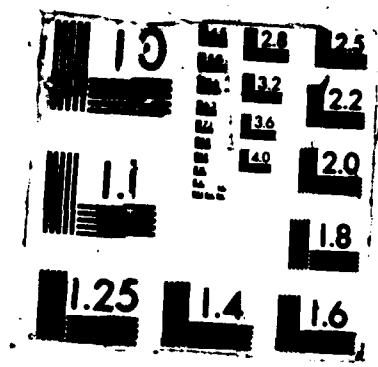
1/1

UNCLASSIFIED

F/G 12/3

NL





12

ESTIMATION IN PARAMETRIC MIXTURE FAMILIES

BY

ALAN E. GELFAND

TECHNICAL REPORT NO. 395

AUGUST 18, 1987

DTIC  
ELECTE  
AUG 28 1987  
S D

Prepared Under Contract  
N00014-86-K-0156 (NR-042-267)  
For the Office of Naval Research

Herbert Solomon, Project Director

Reproduction in Whole or in Part is Permitted  
for any purpose of the United States Government

Approved for public release; distribution unlimited.

DEPARTMENT OF STATISTICS  
STANFORD UNIVERSITY  
STANFORD, CALIFORNIA



Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
DTIC	Avail and/or Special
A-1	

87 8 27 004

# ESTIMATION IN PARAMETRIC MIXTURE FAMILIES

by

Alan E. Gelfand

## 1. Introduction

For mixture distributions of the form

$$f_{\theta}(z) = \int f(z|\eta) dF_{\theta}(\eta) \quad (1)$$

we consider estimation of  $g(\theta)$  under squared error loss. Conditions for the identifiability of  $f_{\theta}(z)$  in  $F_{\theta}(\eta)$  with respect to  $f(z|\eta)$  are discussed in Teicher (1961).

We look at three different problems:

(i) In Section 2 we investigate the possibility of uniformly improving upon an unbiased estimator of  $g(\theta)$ ;

(ii) In Section 3 we offer characterizations of Bayes rules for  $g(\theta)$  as well as a simple complete class theorem;

(iii) In Section 4 we investigate the performance of empirical Bayes rules generated through the EM algorithm. Results will generally be given for  $z, \eta, \theta$  univariate although multidimensional extensions are available in some cases.

Our primary illustrative examples will be in the context of the noncentral chi-square distribution where  $f(z|\eta)$  is  $\chi^2_{p+2\eta}$  and  $F_{\theta}(\eta)$  is Poisson with intensity parameter  $\theta$ . Motivation is provided by recognizing that  $Z$  is inadmissible for estimating  $g(\theta) = E_{\theta}(Z)$  and that  $Z^{-1}$  is inadmissible for

estimating  $g(\theta) = E_{\theta}(Z^{-1})$ . For the former  $\delta(Z) = \max(Z, p)$  clearly dominates  $Z$ . For the latter  $\delta(Z) = \min(Z^{-1}, (p-2)^{-1})$  clearly dominates  $Z^{-1}$ .

## 2. Improving Upon Unbiased Estimators

Let  $T(Z)$  be an unbiased estimator of  $g(\theta)$  and let  $a(\eta) = E(T(Z)|\eta)$ . The "associated conditional problem" is to estimate  $a(\eta)$  under squared error loss within the family  $f(z|\eta)$ . We have the following result.

Theorem 1:  $T(Z) + c \phi(Z)$ ,  $c > 0$ , dominates  $T(Z)$  under squared error loss if

$$\text{cov}_{\theta}(a(\eta), E(\phi|\eta)) \leq 0 \quad \forall \theta \quad (2)$$

$$\beta_{\phi} \equiv \sup_{\eta} \frac{\text{cov}(T, \phi|\eta)}{E(\phi^2|\eta)} \leq -c/2 \quad (3)$$

Proof. By direct calculation we may show that the difference in risk between  $T$  and  $T + c\phi$  is

$$-E_{\theta}(I_{\phi}(\eta)) - 2c \text{cov}_{\theta}(a(\eta), \phi(Z))$$

where

$$I_{\phi}(\eta) = c^2 E(\phi^2|\eta) + 2c E[(T-a)\phi|\eta].$$

But (2) is equivalent to  $\text{cov}_{\theta}(a(\eta), \phi(Z)) \leq 0$  while (3) implies  $I_{\phi}(\eta) \leq 0 \quad \forall \eta$  whence  $T + c\phi$  dominates  $T$ .  $\square$

Remark 1: To hope to satisfy condition (3), we require  $\text{cov}(T, \phi|\eta) \leq 0$  whence it is convenient to choose  $\phi$  inversely related to  $T$ . In many well-known examples, such a choice of  $\phi$  leads to a dominating estimator, e.g., if  $g(\theta) \geq 0$ ,  $T^+$  dominates  $T$  and  $\phi = T^+ - T$  is decreasing in  $T$ .

Remark 2:  $\phi$  decreasing in  $T$  does not necessarily imply condition (2) is met. However, if  $T|\eta$  is a natural exponential family (Morris, 1982), i.e.,  $f(t|\eta) = e^{t\eta - \rho(\eta)}$ , then  $a(\eta) = E(T|\eta) = \rho'(\eta)$  increases in  $\eta$  while for  $E(\phi|\eta)$ ,  $\frac{dE(\phi|\eta)}{d\eta} = \text{cov}(T, \phi|\eta) \leq 0$  implies  $E(\phi|\eta)$  decreases in  $\eta$  so that (2) holds.

Remark 3: If  $T(Z)$  is admissible for  $a(\eta)$  in the conditional problem, it may or may not be admissible for  $g(\theta)$ . For example,  $Z|\eta \sim N(\eta, 1)$  and  $\eta|\theta \sim N(\theta, 1)$ , then  $Z$  is admissible for both  $a(\eta) = \eta$  and  $g(\theta) = \theta$ . If  $Z|\eta \sim \eta^{-1}e^{-z\eta^{-1}}$  and  $\eta|\theta \sim \theta^{-1}e^{-\eta\theta^{-1}}$ ,  $Z/2$  is admissible for  $a(\eta) = \eta$  but  $cZ$ ,  $0 < c < 1/2$  dominates  $Z/2$  for  $g(\theta) = \theta$ .

Remark 4: If  $T$ , unbiased for  $a(\eta)$ , is dominated by  $S$  in the conditional problem, it is possible that  $T$  dominates  $S$  in the unconditional problem. In particular, if we take  $\phi = S - T$ ,  $c = 1$ , we must have (3) hold, i.e.,  $\text{cov}(T, \phi|\eta) < 0$ ,  $\forall \eta$  while the left-hand side of (2) is sufficiently positive  $\forall \theta$ . Examples can readily be constructed using 3 point distributions for  $Z|\eta$ .

Remark 5: In the preceding remark,  $S$  will dominate  $T$  in the unconditional problem if  $T|\eta$  is a natural exponential family using Remark 2.

Suppose instead  $F_\theta(\eta)$  is a natural exponential family in  $\eta$  dominated by  $\mu$  with density  $f(\eta|\theta) = e^{\eta\theta - \chi(\theta)}$  and without loss of generality suppose  $a(\eta) = \eta$ . In this setting, Karlin (1958) supplies conditions such that  $c\eta$  is admissible for  $\chi'(\theta)$  under squared error loss. These conditions require  $c > 0$  and

$$\int_{\theta_0}^{\bar{\theta}} e^{\frac{1-c}{c} \chi(\theta)} d\theta = \infty, \quad \int_{\underline{\theta}}^{\theta_0} e^{\frac{1-c}{c} \chi(\theta)} d\theta = \infty \quad (4)$$

where  $(\underline{\theta}, \bar{\theta})$  is the natural parameter space for  $f(n|\theta)$  and  $\theta_0$  is an arbitrary interior point.

Suppose  $c\eta$  is admissible for  $\chi'(\theta)$ . Is  $cT$  admissible for  $\chi'(\theta)$ ? Remark 3 shows that this is not necessarily the case. We can show that, if Karlin's conditions hold for  $c$

Theorem 2: If

$$E_{\theta} \text{var}(S|\eta) \geq c^2 E_{\theta} \text{var}(T|\eta) \quad \forall \theta \quad (5)$$

then  $S$  cannot dominate  $cT$  in estimating  $\chi'(\theta)$ .

Proof. The proof essentially imitates Karlin's argument.

Suppose  $S$  dominates  $cT$ . Let  $b_S(\theta) = E_{\theta}(S) - cE_{\theta}(T) = E_{\theta}(S) - c\chi'(\theta)$  whence  $b'_S(\theta) = \text{cov}_{\theta}(\eta, E(S|\eta)) - c\chi''(\theta)$ . Therefore

$$(b'_S(\theta) + c\chi''(\theta))^2 \leq \chi''(\theta) \text{var}_{\theta} E(S|\eta)$$

or

$$\frac{(b'_S(\theta) + c\chi''(\theta))^2}{\chi''(\theta)} + E_{\theta} \text{var}(S|\eta) \leq \text{var}_{\theta}(S)$$

and finally

$$E_{\theta}(S - \chi'(\theta))^2 \geq \frac{(b'_S(\theta) + c\chi''(\theta))^2}{\chi''(\theta)} + E_{\theta} \text{var}(S|\eta) + (b_S(\theta) + (c-1)\chi'(\theta))^2.$$

By our supposition the left-hand side of this inequality is at most

$$\begin{aligned}
E_{\theta}(cT - \chi'(\theta))^2 &= c^2 \text{var}_{\theta}(T) + [(c-1)\chi'(\theta)]^2 \\
&= c^2 E_{\theta} \text{var}(T|\eta) + c^2 \chi''(\theta) + [(c-1)\chi'(\theta)]^2.
\end{aligned}$$

Using (5) we obtain

$$\begin{aligned}
&\frac{(b'_S(\theta) + c\chi''(\theta))^2}{\chi''(\theta)} + (b_S(\theta) + (c-1)\chi'(\theta))^2 \\
&\leq c^2 \chi''(\theta) + [(c-1)\chi'(\theta)]^2 \quad \forall \theta.
\end{aligned} \tag{6}$$

Expression (6) is equivalent to Karlin, p. 413, expression (7).

The conditions (4) then imply  $b_S(\theta) \equiv 0$  and  $S = cT$ .  $\square$

Remark 6: This result shows that  $c'T + b$  can't dominate  $cT$  if  $c' > c$ .

Remark 7: If  $T$  is MVUE for  $\eta$ , it is MVUE for  $\chi'(\theta)$ .

Noncentral distributions offer a convenient family of mixtures to study in terms of applying Theorem 1. Gelfand (1983) provides many examples. For the noncentral chi-squared distribution, it is shown that  $\phi$  of the form  $Z^{\beta}$  or of the form  $e^{\beta Z}$  can be used to dominate  $Z$  in estimating  $E_{\theta}(Z)$  and  $Z^{-1}$  in estimating  $E_{\theta}(Z^{-1})$ . Thus  $(Z-p)/2$ , the MVUE of  $\theta$ , can be dominated by estimators of the form  $(Z-p)/2 + cZ^{\beta}$  and of the form  $(Z-p)/2 + ce^{\beta Z}$  for appropriate  $c$  and  $\beta$ . This generalizes earlier results of Perlman and Rasmussen (1975) and of Neff and Strawderman (1976). Of course  $(Z-p)^+/2$  dominates the MVUE as well and has been supported by Chow and Hwang (1983) who argue that it is a simple estimator which "cannot be improved upon uniformly and significantly" and by Saxena and Alam (1982) who

show that it dominates the MLE for  $\theta$ . We will return to this estimator in Section 4.

This section provides a method for uniformly improving upon unbiased estimators in a mixture distribution framework. Dominating biased estimators is a more difficult problem.

### 3. Bayes Estimation

Suppose we let  $\tau_Y(\theta)$  be a family of prior distributions for  $\theta \in \Theta$ . Under squared error loss the generalized Bayes estimator for  $g(\theta)$  is

$$\delta_Y(z) = (f_Y(z))^{-1} \int g(\theta) f_\theta(z) d\tau_Y(\theta) \quad (7)$$

where  $f_Y(z) = \int f_\theta(z) d\tau_Y(\theta)$  is the marginal distribution of  $Z$ .

If the support of  $\tau_Y$  is  $\Theta$  and if  $E_Y g^2(\theta) < \infty$ , then the Bayes risk of  $\delta_Y$  is finite and  $\delta_Y$  is admissible.

Let  $F_\theta(n)$  in (1) be dominated by  $\mu$  with RN derivative  $f(n|\theta)$ . Then

$$\begin{aligned} \delta_Y(z) &= \frac{\iint g(\theta) f(z|\eta) f(\eta|\theta) d\mu(\eta) d\tau_Y(\theta)}{\iint f(z|\eta) f(\eta|\theta) d\mu(\eta) d\tau_Y(\theta)} \\ &= \frac{\int b_Y(\eta) f(z|\eta) d\pi_Y(\eta)}{\int f(z|\eta) d\pi_Y(\eta)} \end{aligned} \quad (8)$$

where  $\pi_Y(\eta)$  is the prior distribution induced on  $\eta$  by  $\tau_Y$ , i.e.,  $d\pi_Y(\eta) = h_Y(\eta) d\mu(\eta)$  and  $h_Y(\eta) = \int f(\eta|\theta) d\tau_Y(\theta)$ , with

$$b_Y(\eta) = h_Y^{-1}(\eta) \int g(\theta) f(\eta|\theta) d\tau_Y(\theta), \quad (9)$$

i.e.,  $b_Y(\eta) = E_Y(g(\theta)|\eta)$ .

Expression (8) shows that

$$\delta_Y(Z) = E(g(\theta)|Z) = E(b_Y(\eta)|Z), \quad (10)$$

i.e., we can calculate the Generalized Bayes rule through the conditional problem. If we let  $c_Y(\eta) = E_Y(g^2(\theta)|\eta)$ , then  $\delta_Y(Z)$  has finite Bayes risk if  $Ec_Y(\eta) < \infty$ .

Let  $F(\eta|\theta)$  be a natural exponential family as in the previous section. If  $g(\theta) = \theta^r$ , then  $b_Y(\eta) = h_Y^{-1}(\eta) \frac{d^r h_Y(\eta)}{d\eta^r}$ ; if  $g(\theta) = e^{r\theta}$ ,  $b_Y(\eta) = h_Y^{-1}(\eta) h_Y(\eta+r)$ .

Gelfand (1983) again offers examples using noncentral distributions. For  $Z$  distributed noncentral chi-squared as in the introduction,

$$h_Y(\eta) = \frac{\int \theta^\eta e^{-\theta}}{\eta!} d\tau_Y(\theta)$$

and

$$\delta_Y(Z) = \frac{\sum b_Y(\eta) (Z/2)^\eta h_Y(\eta) [\Gamma(p/2 + \eta)]^{-1}}{\sum \frac{(Z/2)^\eta h_Y(\eta) [\Gamma(p/2 + \eta)]^{-1}}{\eta}}. \quad (11)$$

If  $g(\theta) = \theta^r$ ,  $b_Y(\eta) = (\eta+r)_r h_Y^{-1}(\eta) h_Y(\eta+r)$ . ( $(x)_y$  denotes the falling factorial of  $y$  terms starting at  $x$ .) At, for example,  $r = 1$  (11) becomes

$$\delta_Y(Z) = 2Z \frac{dJ_Y^2(Z)}{dZ^2} J_Y^{-1}(Z) + \frac{dJ_Y(Z)}{dZ} J_Y^{-1}(Z) \quad (12)$$

where  $J_Y(Z)$  is the denominator of (11). Expression (12) characterizes all generalized Bayes estimates of  $\theta$ . Gelfand notes that setting  $\delta_Y(Z) = (Z-p)/2$ , the MVUE of  $\theta$ , in (12), yields a second order homogeneous linear differential equation

which is not solvable, i.e., the inadmissible MVUE can't be generalized Bayes.

A convenient family for  $\tau_Y(\theta)$  are the distributions Gamma  $(p/2 + v, \gamma)$ ,  $v > -(p-2)/2$  (i.e.,  $E(\theta) = \gamma^{-1}(p/2 + v)$ ) which, for  $g(\theta) = \theta$ , leads to

$$\delta_{\gamma, v}(Z) = \frac{1}{\gamma+1} (p/2 + v + Z \frac{dJ_Y(Z)}{dZ} J_Y^{-1}(Z)) . \quad (13)$$

At  $v = 0$  we obtain the closed form  $\delta_{\gamma, 0}(Z) = (\gamma+1)^{-1}((\gamma+1)^{-1}Z+p)/2$  including the generalized Bayes solution,  $\delta_{0, 0}(Z) = (Z+p)/2$ . These estimators are discussed in Perlman and Rasmussen (1975) and in Saxena and Alam (1982) who note that  $\delta_{0, 0}(Z)$  is dominated by the MVUE. It is straightforward to show that  $\delta_{\gamma, v}(Z)$  increases in  $v$ .

Returning to the rules in (7), we ask if they form a complete class. The conditional problem is not useful here because in the representation of  $\delta_Y$  in (8),  $b_Y$  depends upon the particular prior,  $\tau_Y$ . We attack the problem directly utilizing results of Sacks (1963). His Remark 3, p. 766, argues that with  $g(\theta) = \theta$  the class (7) will be complete under squared error loss if  $f_\theta(z)$  is continuous in both  $\theta$  and  $z$  and, assuming  $-\infty < \theta < \infty$ , for each  $\epsilon > 0$

$$\begin{aligned} \sup_{\theta \leq 0} \frac{\theta^2 f_\theta(z+\epsilon)}{f_\theta(z)} < \infty, \quad \sup_{\theta \geq 0} \frac{\theta^2 f_\theta(z-\epsilon)}{f_\theta(z)} < \infty \\ \lim_{d \rightarrow \infty} \sup_{\theta \geq d} \frac{\theta^2 f_\theta(z-\epsilon)}{f_\theta(z)} = \lim_{d \rightarrow \infty} \sup_{\theta \leq -d} \frac{\theta^2 f_\theta(z+\epsilon)}{f_\theta(z)} = 0 . \end{aligned} \quad (14)$$

Condition (14) will be satisfied if  $f_{\theta}(z+\epsilon)f_{\theta}^{-1}(z) = o(\theta^{-2})$   $\forall z$  and  $\epsilon, -\infty < \epsilon < \infty$ . As noted by Sacks, parts of (14) are not needed if  $\theta$  belongs to a subset of  $R^1$ . We have the following result.

Theorem 3: If  $F_{\theta}(\eta)$ ,  $-\infty < \eta < \infty$ , is a natural exponential family dominated by  $\mu$ , a translation invariant measure, and if  $f(z|\eta) = f(z-\eta)$ , i.e., a translation family, then (7) with  $g(\theta) = \theta$  is a complete class for estimating  $\theta$  under squared error loss.

Proof. By assumptions

$$\frac{\theta^2 f_{\theta}(z+\epsilon)}{f_{\theta}(z)} = \frac{\theta^2 \int f(z + \epsilon - \eta) e^{\eta\theta} d\mu(\eta)}{\int f(z - \eta) e^{\eta\theta} d\mu(\eta)} = \theta^2 e^{\theta\epsilon}$$

from which the conditions in (14) are immediately satisfied.  $\square$

#### 4. Parametric Empirical Bayes Estimation

In the parametric empirical Bayes approach,  $\gamma$  in (7) is assumed unknown and is estimated from the data using the marginal distribution of  $Z$ . If  $\hat{\gamma}$  estimates  $\gamma$  the resultant empirical Bayes estimator is  $\delta_{\hat{\gamma}}(z)$ . The maximum likelihood estimator (MLE) is a frequently employed choice of  $\hat{\gamma}$ . In our setting this requires a very unappealing numerical maximization of a double integral. This would typically be accomplished by algorithms of a Newton or quasi-Newton type. Such algorithms do not guarantee to increase the likelihood at successive iterations. The EM algorithm as described in Dempster et al. (1977) offers an attractive alternative.

To fix notation, let  $t_\gamma$  be dominated by  $\omega$  with RN derivative  $t_\gamma(\theta)$ . At stage  $k$  the algorithm calculates an expectation  $Q(\gamma, \gamma_k) = E_{\gamma_k} \log(f_\gamma(Z, \theta)|z)$  and then maximizes  $Q(\gamma, \gamma_k)$  over all  $\gamma$ . In our case, this simplifies to

$$\max_{\gamma} \int f_{\gamma_k}(\theta|z) \cdot \log t_\gamma(\theta) d\omega(\theta) .$$

This new  $\gamma$  is denoted by  $\gamma_{k+1}$  and the algorithm is repeated until stability is achieved. Such a procedure by its definition may be shown to increase the likelihood with successive iteration (see Dempster et al., '1977). Wu (1983) shows that under minimal assumptions such a procedure yields a stationary value for  $f_\gamma(z)$ . He recommends several EM iterations be tried with different starting  $\gamma$  representative of the parameter space to try to identify local and hopefully a global maximum. Redner and Walker (1984) extensively discuss the use of the EM algorithm for maximum likelihood estimation in mixture distributions. Their focus, however, is on the MLE for  $\theta$  in  $f_\theta(z)$  as in (1) with  $f_\theta(z)$  being a finite mixture density and  $\theta$  a vector including parameters of the distributions being mixed.

If  $t_\gamma$  is an exponential family, i.e.,  $t_\gamma = c(\gamma)e^{\gamma q(\theta)}$ , the algorithm simplifies to maximizing  $\log c(\gamma) + \gamma q_k$  where  $q_k = E_{\gamma_k}(q(\theta)|z)$ . Then  $\gamma_{k+1}$  is a solution to  $-c'(\gamma)c^{-1}(\gamma) = q_k$ . In fact,  $q_{k+1} = E_{\gamma(q_k)}(q(\theta)|z)$  which reduces the algorithm to a stationary or fixed point problem. The conditional representation of  $E_\gamma(q(\theta)|z)$  as  $E_\gamma(b_\gamma(\eta)|z)$  noted in (8) is useful here in

reducing the computation needed for the repeated calculation of the expectation required by the algorithm.

As an example we return to the noncentral chi-squared case under the assumption leading to (13) to obtain the empirical Bayes estimate  $\delta_{\hat{\gamma}, \nu}(Z)$ . It is clear that direct calculation of the MLE,  $\hat{\gamma}$ , is difficult. However, since  $q(\theta) = -\theta'$ , expression (13) up to a sign change sets,  $q_k = \theta_k$  for a given  $\gamma_k$ . But since  $E_{\gamma}(\theta) = \gamma^{-1}(p/2 + \nu)$ , we have  $\gamma_{k+1} \equiv \gamma(q_k) = \theta_k^{-1}(p/2 + \nu)$ .

Writing this explicitly as a fixed point problem, we have

$$\theta = \frac{\theta}{\theta + p/2 + \nu} \left\{ p/2 + \nu + \frac{\sum \left( \frac{z\theta}{2(\theta + p/2 + \nu)} \right)^n \cdot n l_{\nu}(n)}{\sum \left( \frac{z\theta}{2(\theta + p/2 + \nu)} \right)^n \cdot l_{\nu}(n)} \right\} \quad (15)$$

where  $l_{\nu}(n) = \Gamma(p/2 + n + \nu) [\Gamma(p/2 + n) \cdot n!]^{-1}$ . Let the right-hand side of (14) be denoted by  $W_{\nu}(\theta; z)$ . We may show that

- (i)  $W_{\nu}(0; z) = 0$ , i.e., for any  $z$ , 0 is a fixed point;
- (ii)  $W_{\nu}(\theta; z)$  increases in  $\theta$  for fixed  $z$ ;
- (iii)  $W_{\nu}(\theta; z)$  is bounded for fixed  $z$ ;
- (iv) For  $\theta$  small  $W_{\nu}(\theta; z) \approx \theta$  regardless of  $z$ ;
- (v)  $W_{\nu}(\theta; z)$  has at most one positive fixed point for a fixed  $z$ .

Hence, given  $\nu$  and  $z$ , a plot of  $W_{\nu}(\theta; z)$  vs.  $\theta$  assumes one of the two forms in Figure 1.

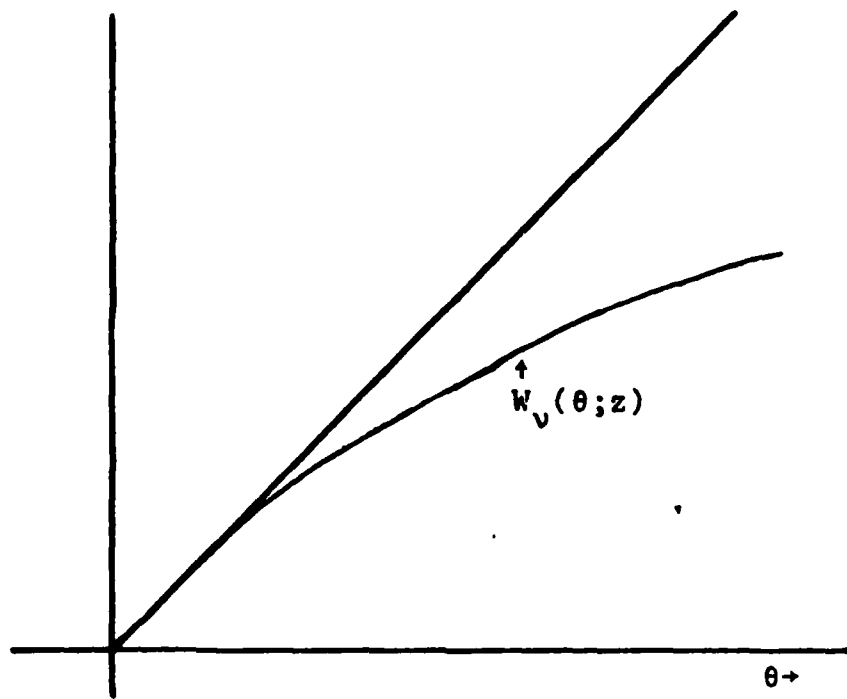


FIGURE 1(a)

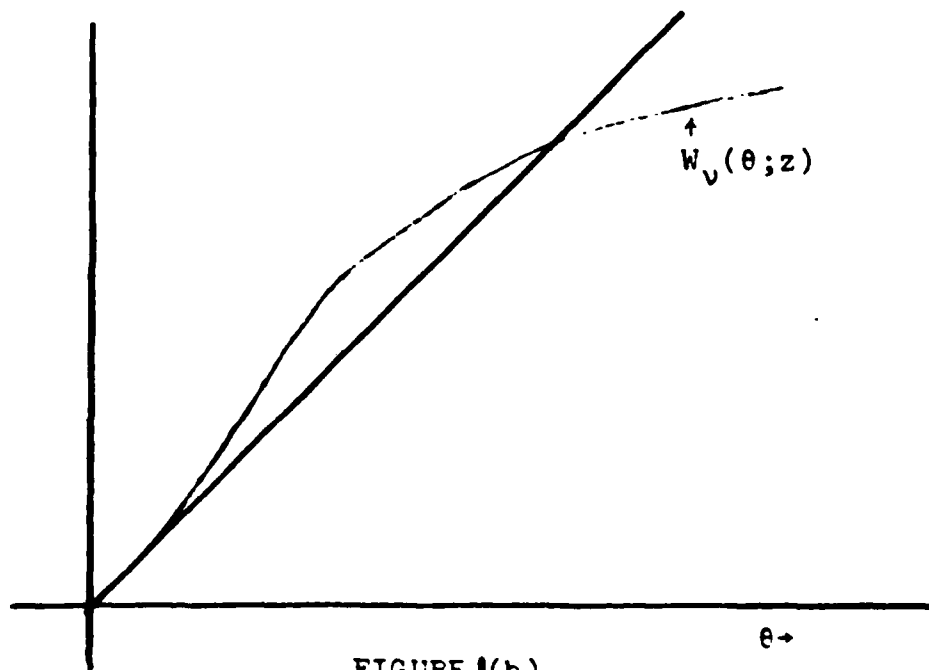


FIGURE 1(b)

With increasing  $z$  the plot will change from Figure/(a) to Figure/(b). As a result, if we employ the EM algorithm, we will find that for any starting  $\theta$  under Figure (a) and for starting  $\theta$  sufficiently small under (b),  $\theta_k \rightarrow 0$ , i.e.,  $\gamma_k \rightarrow \infty$ . Moreover, because of (iv), the convergence will be extremely slow (e.g., 3,000 iterations may yield  $\theta_k$  of the order of  $10^{-2}$ ). It is noteworthy that when  $\theta_k \rightarrow 0$ , i.e.,  $\hat{\theta} = 0$  ( $\hat{\gamma} = \infty$ ) we, in fact, minimize  $f_{\gamma}(z)$ , i.e., the EM algorithm will fail to maximize the likelihood. But in terms of empirical Bayes estimation, from (13),  $\delta_{\infty, \gamma}(z) = 0$  is a reasonable guess for  $\theta$  if  $z$  is sufficiently small. If there is a fixed point  $\theta > 0$ , the corresponding  $\gamma$  must be the MLE. In implementing the EM algorithm, we should begin with  $\gamma$  small, i.e.,  $\theta$  large to insure finding this fixed point if it exists. Moreover, if after, say 300 iterations  $\theta_{300}$  is small (say  $10^{-2}$ ) and decreasing, we will conclude that  $\hat{\theta}_k \rightarrow 0$ . (It is possible that the nonzero fixed point lies below  $\theta_{300}$ , but practically this is of little concern.) The maddeningly slow convergence of the algorithm even to the unique MLE may make the following alternative attractive. Using a rough plot of  $W_{\gamma}(\theta; z)$  versus  $\theta$  for a few choices of  $z$  should enable, when  $z$  is sufficiently large, identification of an appropriate initial  $\theta$  to insure convergence of  $\gamma_k$  to the MLE or to conclude that  $z$  is sufficiently small so that  $\theta_k \rightarrow 0$ . We note that for any  $z$  and any starting  $\gamma$  the empirical Bayes estimator resulting from this algorithm will be  $\geq 0$  and, in fact,

if  $v = 0$ , the right-hand side of (13) simplifies and

$\delta_{\gamma,0}^{\wedge}(Z) = (Z-p)^{+}/2$  (in fact,  $\hat{\gamma} = [p(Z-p)^{-1}]^{+}$ ), i.e., the positive part version is empirical Bayes. Perlman and Rasmussen (1975) observed that the MVUE is itself empirical Bayes if we take  $\delta_{\gamma,0}^{\wedge}(Z)$  with  $\hat{\gamma} = p(Z-p)^{-1}$ .

An extensive simulation was conducted to study the risk behavior of the  $\delta_{\gamma,v}^{\wedge}(z)$ . Cases  $p = 6$  with  $v = -2, -1, 0, 1, 2, 5, 10$  and  $p = 12$  with  $v = -4, -2, 0, 2, 4, 10, 20$  were examined using 5,000 replications. The algorithm was allowed 300 iterations on each replication. Convergence was declared if  $|\gamma_{k+1} - \gamma_k| < 10^{-4}$ . If for any  $k$ ,  $e_k < 10^{-2}$ ,  $\theta = 0$  was taken as the estimate. If the algorithm failed to converge after 300 iterations,  $\theta_{300}$  was taken as the estimate. Starting points of (i)  $\gamma_0 = 1$  and (ii)  $\gamma_0 = (Z-p)^{-1}(p+2v)$  if  $Z > p$ ,  $\gamma_0 = 1$  if  $Z < p$ , were tried. The choice  $\gamma_0 = 1$  may be viewed as the "center" of the parameter space in that it corresponds, for the induced negative binomial prior on  $\eta$ , to a success probability of .5. The choice  $\gamma_0 = (Z-p)^{-1}(p+2v)$  arises from the fact that  $E_{\gamma}(Z) = p + \gamma^{-1}(p+2v)$ . Both starting values were successful in obtaining the unique MLE. However, (ii) tended to converge more quickly. Convergence tended to be slower with increasing  $v$  although more frequently to the unique MLE than to the fixed point at 0, i.e., more frequently we would be in the case of Figure 1(b). Increasing  $p$  from 6 to 12 led to quicker convergence again more frequently to the MLE.

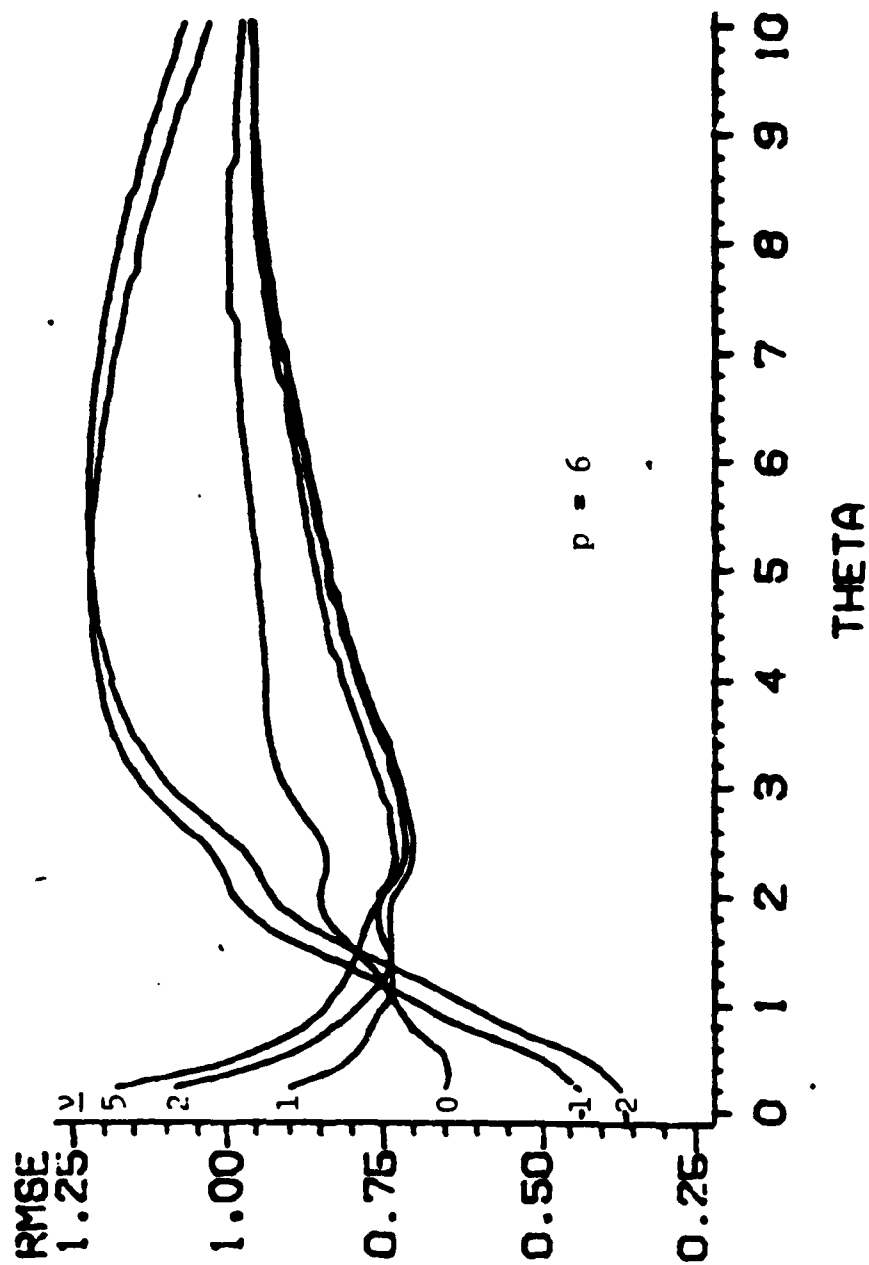


FIGURE 2: RMSE Behavior,  $p = 6$ , for  $\delta_{\hat{\gamma}, v}$ ,  $v = -2, -1, 0, 1, 2, 5$

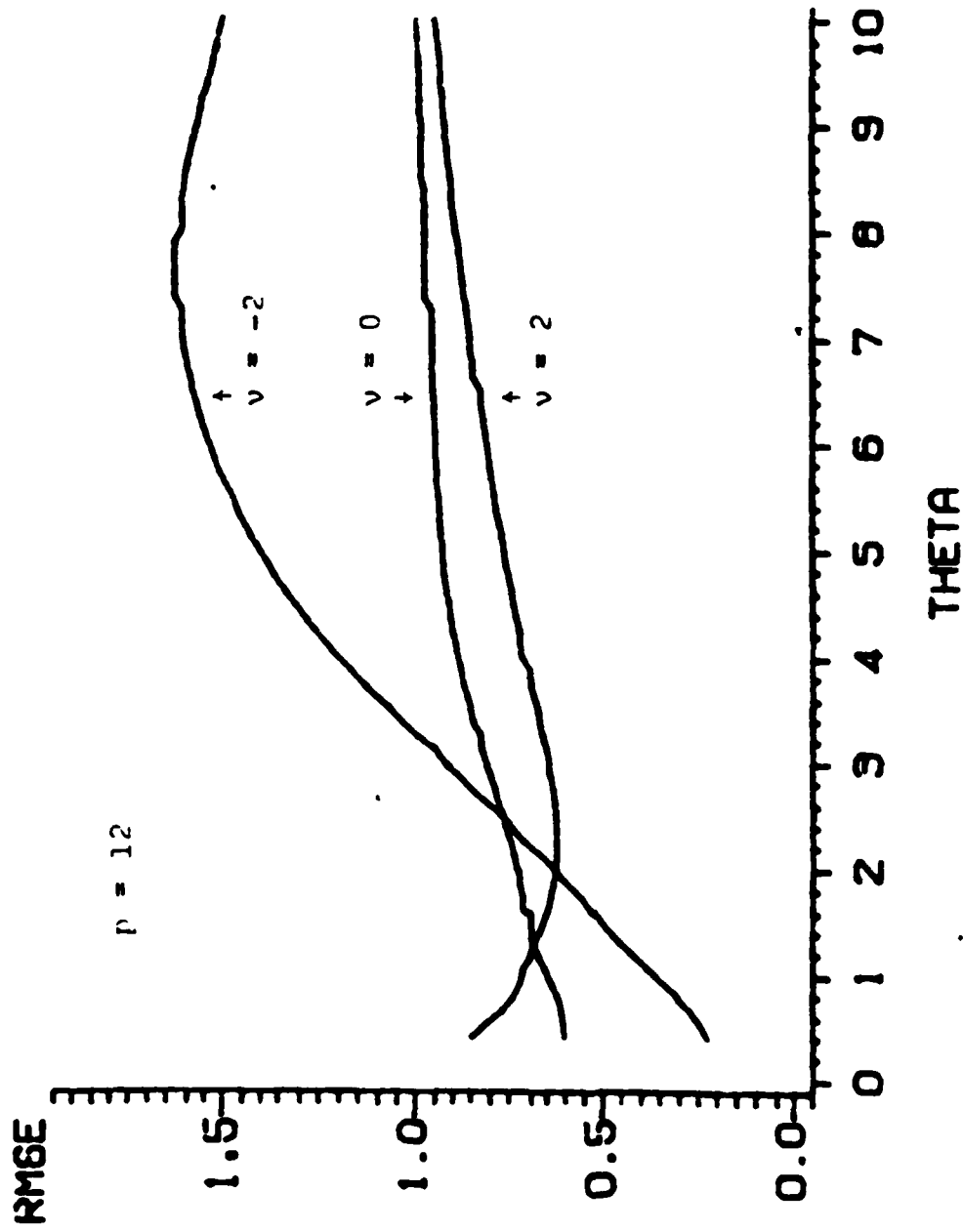


FIGURE 3: RMSE Behavior,  $p = 12$ , for  $\delta_{\gamma, v}$ ,  $v = -2, 0, 2$

Figures 2 and 3 display the results for  $p = 6$  and  $p = 12$ , respectively, in terms of relative mean square error (RMSE), i.e., the MSE of  $\delta_{\hat{Y},v}$  relative to  $p/2 + 2\theta$ , that of the MVUE. In Figure 1, we present 6 values of  $v$  surrounding  $v = 0$  which we recall yields  $(2-p)^+/2$ . In Figure 2, we simplify to  $v = -2, 0, 2$ . It is noteworthy that the  $v = 1, 2$  estimates not previously discussed in the literature dominate the positive part MVUE except for  $\theta$  small.

#### Acknowledgment

The author thanks Steven Leeds for performing the simulation study.

#### References

- Chow, M.S. and Hwang, J.T. (1983). The Comparison of Estimators for the Noncentrality of a Chi-square Distribution. Tech. Report, Department of Mathematics, Cornell University.
- Dempster, A., Laird, N., and Rubin, D. (1977). Maximum Likelihood from Incomplete Data Via the EM Algorithm. J. Royal Statist. Soc. B39, 1-38.
- Gelfand, A. (1983). Estimation in Noncentral Distributions. Commun. Statist. 12 (4), 463-475.
- Karlin, S. (1958). Admissibility for Estimation with Quadratic Loss. Ann. Math. Statist. 29, 406-436.
- Morris, C. (1982). Natural Exponential Families with Quadratic Variance Functions. Ann. Statist. 10, 65-80.
- Neff, N. and Strawderman, W.E. (1976). Further Remarks on Estimating the Parameter of a Noncentral Chi-squared Distribution. Commun. Statist. 5, 65-67.

- Ferlman, M. and Rasmussen, U. (1975). Some Remarks on Estimating a Noncentrality Parameter. Commun. Statist. 4, 455-468.
- Redner, R. and Walker, H. (1984). Mixture Densities, Maximum Likelihood and the EM Algorithm. SIAM Review 26, 195-239.
- Sacks, J. (1963). Generalized Bayes Solutions in Estimation Problems. Ann. Math. Statist. 34, 751-768.
- Saxena, K.M.L. and Alam, K. (1982). Estimation of the Noncentrality Parameter of a Noncentral Chi-squared Distribution. Ann. Statist. 10 (3), 1012-1016.
- Teicher, H. (1961). Identifiability of Mixtures. Ann. Math. Statist. 32 (1), 244-248.
- Wu, C.F.J. (1983). On the Convergence Properties of the EM Algorithm. Ann. Statist. 11 (1), 95-103.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 395	2. GOVT ACCESSION NO. <b>AD-A183818</b>	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Estimation In Parametric Mixture Families		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Alan E. Gelfand		8. CONTRACT OR GRANT NUMBER(s) N00014-86-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-267
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Statistics & Probability Program Code 1111		12. REPORT DATE August 18, 1987
		13. NUMBER OF PAGES 21
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report)  APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)  Parametric Mixture Distributions, Bayes Rules, Empirical Bayes Rules, EM Algorithm, Non-Central Chi-square Distribution		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number)  PLEASE SEE FOLLOWING PAGE.		

DD FORM 1473  
1 JAN 73EDITION OF 1 NOV 65 IS OBSOLETE  
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

TECHNICAL REPORT NO. 395

## 20. ABSTRACT

For parametric mixture distributions indexed by say a univariate parameter  $\theta$  we investigate estimation of  $g(\theta)$  under squared error loss. First we propose a method for uniformly improving upon an unbiased estimator of  $g(\theta)$ . Second we characterize Bayes estimators of  $g(\theta)$  and give a simple complete class theorem. Finally we study the performance of empirical Bayes rules generated using the EM algorithm. Application in the context of the noncentral chi-square distribution provides examples.

END

9-87

Dtic